UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/944,919	08/31/2001	Maria Castellanos	HP-10007912	3894

7590   03/04/2004

HEWLETT-PACKARD COMPANY
Intellectual Prperty Administration
P.O. Box 272400
Fort Collins, CO 80527-2400

EXAMINER
AMSBURY, WAYNE P

ART UNIT	PAPER NUMBER
2171	6

DATE MAILED: 03/04/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

Office Action Summary	Application No.	Applicant(s)
	09/944,919	CASTELLANOS ET AL.
	Examiner	Art Unit
	Wayne Amsbury	2171

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.
- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133).
  Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *28 January 2004*.

2a)☒ This action is **FINAL**.   2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-12 and 15-35* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-12 and 15-35* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *8/31/01* is/are: a)☐ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☐ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5)☐ Notice of Informal Patent Application (PTO-152)

6)☐ Other: _____.

CLAIMS 1-12 AND 15-35 ARE PENDING

1. The text of those sections of Title 35, U.S. Code not included in this action can be found in a prior Office action.

2. **Claims 1-12 and 15-35 are rejected under 35 U.S.C. 103(a) as being unpatentable over Domini et al (Domini), US 6,085,206, 4 July 2000 in light of Zamora, US 4,965,763, 23 October 1990.**

Domini is directed to the correction of both spelling and grammatical construction in a document. Misspelling is one form of dirty text; the extraction of grammatical constructions is a form of data mining. Domini does not explicitly address providing a summary of content or the scoring and ranking of sentences. It is well known to index, rank, and sort documents based on components, as evidenced by Zamora [ABSTRACT; SUMMARY; COL 38 lines 8-20 and elsewhere]. Zamora is directed to extracting and analyzing components of documents, particularly sentences, in the form of frames [FIG 3, 18; COL 4 lines 37-41, 50-53 and elsewhere].

Zamora extracts various forms of syntactic and morphological properties [COL 12 lines 42-52], but does not explicitly correct spelling and other forms of dirty text, but the proper determination of names, addresses and the like requires correction of at least typographical errors.

It would have been obvious to one of ordinary skill in the art at the time of the invention to apply the cleanup procedure of Domini prior to the analysis of Zamora because this would provide a more accurate basis for document analysis.

It would have been obvious to one of ordinary skill in the art at the time of the invention to apply the PIE (Parametric Information Extraction) system of Zamora to that of Domini because it would enhance the retrieval of documents [Zamora BACKGROUND; OBJECTS; SUMMARY].

As to **claims 1-2 and 8**, Domini detects misspelled words and allow the user to correct them and provides for the automatic replacement of such words with correct spellings [COL 13 lines 19-42]. This removes an instance of dirty text within a document to produce a cleaned document. The parsing of a sentence [COL 4 lines 10-29] is a data mining operation the cleaned sentence.

The index of Zamora corresponds to a summary, so does the inverted file of FIG 18. The PIE RESULTS [COL 16] themselves correspond to a summary of information is an example format such as the MEMO FORMAT [COL 16]. As noted above, proper and accurate determination on the components of the PIE subsumes that clean text is analyzed.

As to **claims 3-4**, neither Domini nor Zamora explicitly addresses the extraction of computer code or a table, but these are well known components of documents of various types. In particular, web pages include a form of computer code.

It would have been obvious to one of ordinary skill in the art at the time of the invention to include these structures in the mining and analysis of documents because to omit them would severely limit the applicability of the system.

As to **claim 5**, Domini determines the beginning and end of sentences [COL 3 lines 42-54]. Zamora requires the determination of the components of a frame [COL 6 lines 1-16 and elsewhere], which requires determination of at least statements.

As to **claim 6, it would have been obvious** to one of ordinary skill in the art at the time of the invention to rank the business documents of Zamora in terms of significant sentences such as the subject statement and the reference statement [COL 6 lines 1-16] because they are important for determining the significance of a business document.

As to **claim 7**, it is clear that the frames of Zamora involve both narrative and non-narrative text.

As to **claim 9**, the text mining of Domini contains selections of functions to be performed, as noted at COL 4 lines 21-29.

As to **claim 10**, Domini provides for switches that control the mining operations and which inherently involve a parameter, as seen in FIG 3-4.

The elements of **claims 11-12 and 15-30** are rejected in the analysis above and these claims are rejected on that basis.

As to **claim 31**, ranking is implicitly from highest-to-lowest in one direction and lowest-to-highest in the other.

As to **claim 32**, Zamora addresses keyword ranking of documents [COL 2 lines 24-31], but not explicitly individual sentences. However, Zamora also discusses the significance of particular authors, address, subject, and dates [COL 2 lines 42-68]. These components of a business document correspond in a general way to sentences. Furthermore, Zamora is directed to finding user-defined expressions [COL 2 lines 6-14] and subject statements [COL 2 lines 48-52]. To the extent that these do not correspond to sentences it would have been obvious to one of ordinary skill in the art at the time of the invention to apply the ranking to sentences because sentences are an efficient mode of inclusion for an expression, statement, or business letter parameter that can determine the significance of a business document.

As to **claim 33**, the frames and PIEs of Zamora are directly related to location, and determine the significance of document components.

As to **claim 34**, Zamora addresses the significance of semantic elements to the evaluation of queries [COL 2 lines 58-68].

As to **claim 35**, the use of vectors and the cosine of an angle between them to compare text elements is well known in the art, taught within the paradigm of latent semantic indexing, to such an extent that it would have been obvious to one of ordinary skill in the art at the time of the invention to apply this technique to relate text fragments such as sentences because it is efficient to apply a known method.

3. Applicant's arguments filed 1/28/04 have been fully considered but they are not persuasive.

The Applicant states at page 9 of the Response:

*Domini fails to describe **performing a data mining operation on a cleaned document**. The Examiner incorrectly concluded that correction of grammar is a form of data mining. Similar to Applicant, Domini defines the extraction of grammatical constructions as removal of dirty text.*

It is noted that the Specification does not contain the terms *extraction* and *grammatical* in the same sentence. Furthermore, Claim 5 defines the identifying of a sentence as a data mining operation:

[claim] 5. The method for mining a document containing dirty text as recited in Claim 1, wherein said performing a data mining operation further comprises identifying a sentence within said cleaned document by identifying a beginning and an end of said sentence.

Claim 5 clearly implies that data mining includes recognition of sentence structure, which is certainly grammar, and is affected by dirty text. The term "data mining" is defined only by example in the Specification, but certainly includes the example of claim 5.

The Applicant bases arguments on the statement at ¶ 2 page 10 of the Response:

Neither Domini nor Zamora teaches or suggests suggests (sic) a two-step cleaning process. Specifically, neither Domini nor Zamora teaches or suggests a general cleaning followed by a domain and task-specific cleaning of the document including removing instances of computer code and tables.

Firstly, Domini teaches data mining in at least one sense that matches that of the Specification, as discussed above. Secondly, motivation was provided in the rejection for using these references in concert. Thirdly, Applicant is reminded of MPEP 2144:

The rationale to modify or combine the prior art does not have to be expressly stated in the prior art; the rationale may be expressly or impliedly contained in the prior art or it may be reasoned from knowledge generally available to one of ordinary skill in the art, established scientific principles, or legal precedent established by prior case law.

In particular, Zamora depends on basic determination of the elements and semantic content of documents [COL 1 line 47 to Col 2 line 41; COL 4 lines 30-41; and elsewhere]. It is clear that many forms of dirty text would corrupt the process, and need to be done prior to determination of sentence structure. For instance, a misspelled word cannot be recognized as repeated by a correctly spelled word without an initial general cleanup. The statement of the Applicant above repeats the word "suggests" which would skewer results dependent on word counts and grammatical constructions. One of ordinary skill in the art would certainly be aware of the need for general cleanup prior to analysis of the structure of a document, and be reminded of this by frequent use of word processors.

Finally, the rejection recognized that neither Domini nor Zamora explicitly addresses the extraction of computer code or a table, but a motivation was supplied for incorporating these extractions into the combined system.

4. **THIS ACTION IS MADE FINAL.** Applicant is reminded of the extension of

time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE

MONTHS from the mailing date of this action. In the event a first reply is filed within

TWO MONTHS of the mailing date of this final action and the advisory action is not

mailed until after the end of the THREE-MONTH shortened statutory period, then the

shortened statutory period will expire on the date the advisory action is mailed, and any

extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of

the advisory action. In no event, however, will the statutory period for reply expire later

than SIX MONTHS from the mailing date of this final action.

5. Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Wayne Amsbury whose telephone number is 703-305-

3828. The examiner can normally be reached on M-TH 7-5.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Safet Metjahic can be reached on 703-308-1436. The fax phone number for

the organization where this application or proceeding is assigned is 703-872-9306.

Information regarding the status of an application may be obtained from the

Patent Application Information Retrieval (PAIR) system. Status information for

published applications may be obtained from either Private PAIR or Public PAIR.

Status information for unpublished applications is available through Private PAIR only.

For more information about the PAIR system, see http://pair-direct.uspto.gov. Should

you have questions on access to the Private PAIR system, contact the Electronic

Business Center (EBC) at 866-217-9197 (toll-free).

WPA

WAYNE AMSBURY
PRIMARY PATENT EXAMINER